

APPLICATIONS OF A FINITE-DIMENSIONAL DUALITY PRINCIPLE TO SOME PREDICTION PROBLEMS

YUKIO KASAHARA, MOHSEN POURAHMADI AND AKIHIKO INOUE

ABSTRACT. Some of the most important results in prediction theory and time series analysis when finitely many values are removed from or added to its infinite past have been obtained using difficult and diverse techniques ranging from duality in Hilbert spaces of analytic functions (Nakazi, 1984) to linear regression in statistics (Box and Tiao, 1975). We unify these results via a finite-dimensional duality lemma and elementary ideas from the linear algebra. The approach reveals the inherent finite-dimensional character of many difficult prediction problems, the role of duality and biorthogonality for a finite set of random variables. The lemma is particularly useful when the number of missing values is small, like one or two, as in the case of Kolmogorov and Nakazi prediction problems. The stationarity of the underlying process is not a requirement. It opens up the possibility of extending such results to nonstationary processes.

1. INTRODUCTION

Irregular observations, missing values and outliers are common in time series data (Box and Tiao (1975), Brubacher and Wilson (1976)). A framework for dealing with such anomalies is that of $X = \{X_t\}_{t \in \mathbb{Z}}$ being a \mathbb{C} -valued, mean-zero, weakly stationary stochastic process with the autocovariance function $\gamma = \{\gamma_k\}_{k \in \mathbb{Z}}$ and the spectral density function f : $E[X_k \bar{X}_l] = \gamma_{k-l} = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-i(k-l)\lambda} f(\lambda) d\lambda$. Then, the problem can be formulated as that of predicting or approximating an unknown value X_0 based on the observed values $\{X_t; t \in S\}$ for a given index set $S \subset \mathbb{Z} \setminus \{0\}$ and the knowledge of the autocovariance of the process. Such a problem is quite important to applications in business, economics, engineering, physical and natural sciences etc., and belongs to the area of prediction theory of stationary stochastic processes developed by Wiener (1949) and Kolmogorov (1941) (see also Pourahmadi (2001)). By restricting attention to linear predictors and using the least-squares criterion to assess the goodness of predictors, a successful solution seeks to address the following two goals:

- (P₁) Express the linear least-squares predictor of X_0 , denoted by $\hat{X}_0(S)$, and the prediction error $X_0 - \hat{X}_0(S)$ in terms of the observable $\{X_t; t \in S\}$.
- (P₂) Express the prediction error variance $\sigma^2(S) = \sigma^2(f, S) := E|X_0 - \hat{X}_0(S)|^2$ in terms of f .

Date: August 29, 2007.

Key words and phrases. Finite prediction problems, biorthogonality and duality, missing values, stationary time series, Wold decomposition.

2000 *Mathematics Subject Classification* Primary 62M20; Secondary 60G10; 60G25.

The link between solutions of finite and infinite past prediction problems serves as a natural bridge between time series analysis and prediction theory. From the dawn of modern time series analysis, the works of Slutsky and Yule in the 1920's and Wold in the 1930's have been instrumental in achieving the goal (P_1) in the time-domain using the *finite* past. Subsequently, the classes of autoregressive (AR), moving-average (MA) and mixed autoregressive and moving-average (ARMA) models have played major roles in the development of time-domain techniques using the autocovariance function of the process (see Box *et al.* (1994)). Nowadays, these techniques are implemented by solving the Yule–Walker equations via the celebrated Durbin–Levinson algorithm and the innovation algorithm (see Brockwell and Davis (1991)). On the other hand, the spectral-domain techniques in prediction of stationary processes, advocated by Kolmogorov and Wiener in the early 1940's, rely on the spectral representations of the process and its covariance (Kolmogorov (1941), Wiener (1949), Pourahmadi (2001)).

The focus in prediction theory is more on the goal (P_2). The celebrated Szegö–Kolmogorov–Wiener theorem gives the variance of the one-step ahead prediction error based on the *infinite* past indexed by the “half-line” $S_0 := \{\dots, -2, -1\}$ by

$$(1.1) \quad \sigma^2(f, S_0) = \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda\right) > 0$$

if $\log f$ is integrable, and otherwise $\sigma^2(S_0) = 0$. However, when the first n consecutive integers are removed from S_0 or for the index set $S_{-n} := \{\dots, -n-2, -n-1\}$, $n \geq 0$, the formula for the $(n+1)$ -step prediction error variance (Wold (1938), Kolmogorov (1941)) is

$$(1.2) \quad \sigma^2(f, S_{-n}) = |b_0|^2 + |b_1|^2 + \dots + |b_n|^2, \quad n = 0, 1, \dots,$$

where $\{b_j\}$, the MA coefficients of the process, is related to the Fourier coefficients of $\log f$ and $|b_0|^2 = \sigma^2(S_0)$ (see Nakazi and Takahashi (1980) and Pourahmadi (1984); see also Section 3 below).

A result similar to (1.1) for the interpolation of a single missing value corresponding to the index set $S_\infty := \mathbb{Z} \setminus \{0\}$ was obtained by Kolmogorov (1941). Specifically, the interpolation error variance is given by

$$(1.3) \quad \sigma^2(f, S_\infty) = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda)^{-1} d\lambda \right)^{-1} > 0$$

if $f^{-1} \in L^1 := L^1([-\pi, \pi], d\lambda/(2\pi))$, and otherwise $\sigma^2(S_\infty) = 0$. The corresponding prediction problem for the smaller index set $S_n := \{\dots, n-1, n\} \setminus \{0\}$, $n \geq 0$, was stated as open in Rozanov (1967, p. 107) and is perhaps one of the most challenging problems in prediction theory next to (1.1). The index set S_n is, indeed, of special interest as it forms a bridge connecting S_0 and S_∞ ; it reduces to S_0 when $n = 0$ and tends to S_∞ as $n \rightarrow \infty$. In a remarkable paper in 1984, Nakazi using delicate, but complicated analytical techniques (and assuming that $f^{-1} \in L^1$) showed that

$$(1.4) \quad \sigma^2(f, S_n) = (|a_0|^2 + |a_1|^2 + \dots + |a_n|^2)^{-1}, \quad n = 0, 1, \dots,$$

where $\{a_j\}$ is related to the AR parameters of the process (see Section 3 below).

From (1.2) and (1.4), the question naturally arises as why there is such an “inverse-dual” relationship between them. In this regard, it is worth noting that Nakazi’s technique, if interpreted properly, amounts to reducing computation of $\sigma^2(f, S_n)$ to that of the $(n+1)$ -step prediction error variance of another stationary

process $\{Y_t\}$ with the spectral density function f^{-1} which turns out to be the *dual* of $\{X_t\}$ (see Definition 2.1 and Section 3.5). His result and technique have spawned considerable research in this area in the last two decades; see Miamee and Pourahmadi (1988), Miamee (1993), Cheng *et al.* (1998), Frank and Klotz (2002), Klotz and Riedel (2002) and Bondon (2002). A unifying feature of most of the known results thus far seems to be a fundamental duality principle (Cheng *et al.* (1998), Urbanik (2000)) of the form

$$(1.5) \quad \sigma^2(f, S) \cdot \sigma^2(f^{-1}, S^c) = 1,$$

where S^c is the complement of S in $\mathbb{Z} \setminus \{0\}$ and $f^{-1} \in L^1$. The first occurrence of (1.5) seems to be in the 1949 Russian version of Yaglom (1963) for the case of deleting finitely many points from S_∞ . Proof of (1.5), in general, like those of the main results in Nakazi (1984), Miamee and Pourahmadi (1988), Cheng *et al.* (1998), and Urbanik (2000), is long, unintuitive and relies on duality techniques from functional and harmonic analysis and requires $f^{-1} \in L^1$ which is not natural for an index set like S_n . Surprisingly, a version of (1.5) in a rather disguised form was developed in Grenander and Rosenblatt (1954, Theorem 1), as the limit of a quadratic form involving Szegő's orthogonal polynomials on the unit circle, see also Simon (2005, p.165). Unfortunately, it had remained dormant and not used in the context of prediction theory, except in Pourahmadi (1993).

In this paper, we establish a finite-dimensional duality principle (Lemma 2.4), which encapsulates (1.5) in a transparent and useful manner. The concept of dual of a random vector plays a central role as does the Cholesky decomposition of its covariance matrix. We use this duality principle to unify and solve some prediction problems related to removing a finite number of indices from S_n and S_∞ . The outline of the paper is as follows. In Section 2, we present the main lemma, some auxiliary facts about dual of a random vector and their consequences for computing the prediction error variances and predictors. In Section 3, using the lemma we first solve three finite prediction problems for X_0 based on the knowledge of $\{X_t; t \in K\}$ with $K = \{-m, \dots, n\} \setminus (M \cup \{0\})$, $m, n \geq 0$, where M , the index set of the missing values, is relatively small. Then we obtain the solutions of Kolmogorov, Nakazi, and Yaglom's prediction problems in a unified manner by studying the limit of the solutions by letting $m \rightarrow \infty$, followed by $n \rightarrow \infty$. In particular, we find explicit formula for the dual of the process $\{X_t; t \leq n\}$ for a fixed n , which does not seem to be possible using the technique of Urbanik (2000), Klotz and Riedel (2002) and Frank and Klotz (2002). This is useful in developing series representations for predictors and interpolators, and sheds light on the approaches of Bondon (2002) and Salehi (1979). In Section 4, we close the paper with some discussions.

Finally, we should point out that the two simple formulas (1.2) and (1.4) and their extensions provide explicit and informative expressions for the prediction error variances. Like their predecessors (1.1) and (1.3), they serve as yardsticks to assess the impact (worth) of observations in predicting X_0 when they are added to or deleted from the infinite past and highlight the role of the autoregressive and moving-average parameters for this purpose; see Pourahmadi and Soofi (2000). In fact, Bondon (2002, Theorem 3.3; 2005) shows that a finite number of missing values do not affect the prediction of X_0 if and only if the AR parameters corresponding to the indices of those missing values are zero. Furthermore, the examples in Section 3 indicate how the interpolators of the missing values can be computed

rigorously without resorting to formal derivations (Box and Tiao (1975), Brubacher and Wilson (1976) and Budinsky (1989)).

2. A FINITE-DIMENSIONAL DUALITY PRINCIPLE

In this section, an elementary result is stated as a finite-dimensional duality lemma, which we use in Section 3 to solve and unify various challenging prediction problems through the limit of the solutions of their finite past counterparts.

For a finite index set N , let H_N be the class of vectors $X = (X_j)_{j \in N}$ of random variables with zero-mean and finite variance on a probability space (Ω, \mathcal{F}, P) :

$$H_N := \{X = (X_j)_{j \in N}; X_j \in L^2(\Omega, \mathcal{F}, P), E[X_j] = 0, j \in N\}.$$

As usual, we consider the inner product $(Y, Z) := E[Y\bar{Z}]$ and norm $\|Y\| := E[|Y|^2]^{1/2}$ for random variables in $L^2(\Omega, \mathcal{F}, P)$.

Definition 2.1. Let N be a finite index set and $X \in H_N$. A random vector $Y \in H_N$ is called the *dual* of X if it satisfies the following conditions:

- (i) The components $Y_j, j \in N$, belong to $\text{sp}\{X_k; k \in N\}$.
- (ii) X and Y are *biorthogonal*: $(X_i, Y_j) = \delta_{ij}$ for $i, j \in N$, or $\text{Cov}(X, Y) = I$.

For $X \in H_N$, $l \in N$ and $K \subset N$, we write $\hat{X}_l(K)$ for the linear least squares predictor of X_l based on $\{X_k; k \in K\}$, i.e., the orthogonal projection of X_l onto $\text{sp}\{X_k; k \in K\}$. For the sake of completeness and ease of reference, in the next two propositions we summarize the characterization, interpretation and other basic information about the dual of a random vector in terms of its covariance matrix and certain prediction errors.

Proposition 2.2. Let N be a finite index set and $X \in H_N$. Then, the following conditions are equivalent:

- (1) The components $X_j, j \in N$, of X are linearly independent.
- (2) The covariance matrix $\Gamma = (\gamma_{i,j})_{i,j \in N}$ of X with $\gamma_{i,j} = (X_i, X_j)$ is nonsingular.
- (3) X is minimal: $X_j \notin \text{sp}\{X_i; i \in N, i \neq j\}$ for $j \in N$.
- (4) X has a dual.

Proof. Clearly, (1)–(3) are equivalent. Assume (3) and define $Y = (Y_j)_{j \in N} \in H_N$ by $Y_j = (X_j - \hat{X}_j(N_j)) / \|X_j - \hat{X}_j(N_j)\|^2$, where $N_j := N \setminus \{j\}$. Then Y_j belongs to $\text{sp}\{X_k; k \in N\}$, and $(X_i, Y_j) = \delta_{ij}$ holds:

$$(X_j, Y_j) = \frac{(X_j, X_j - \hat{X}_j(N_j))}{\|X_j - \hat{X}_j(N_j)\|^2} = \frac{(X_j - \hat{X}_j(N_j), X_j - \hat{X}_j(N_j))}{\|X_j - \hat{X}_j(N_j)\|^2} = 1,$$

and for $i \neq j$,

$$(X_i, Y_j) = \frac{(X_i, X_j - \hat{X}_j(N_j))}{\|X_j - \hat{X}_j(N_j)\|^2} = 0.$$

Thus Y is a dual of X , and hence (4). Conversely, assume (4) and let Y be a dual of X . If X is not minimal, then there exists $j \in N$ such that $X_j \in \text{sp}\{X_i; i \in N, i \neq j\}$, that is, $X_j = \sum_{i \neq j} c_i X_i$ for some $c_i \in \mathbb{C}$, and, since $(X_i, Y_j) = 0$ for $i \neq j$, we have $(X_j, Y_j) = \sum_{i \neq j} c_i (X_i, Y_j) = 0$. However, this contradicts $(X_j, Y_j) = 1$. Thus, X is minimal, and (3) follows. \square

The proof reveals the importance of the “standardized” interpolation errors of components of X in defining its dual. More explicit representations and other properties of the dual are given next.

Proposition 2.3. *For a finite index set N , let $X \in H_N$ with covariance matrix Γ . Assume that X has a dual Y . Then the following assertions hold:*

- (1) *The dual Y is unique.*
- (2) *The dual Y is given by $Y_j = (X_j - \hat{X}_j(N_j)) / \|X_j - \hat{X}_j(N_j)\|^2$ with $N_j := N \setminus \{j\}$ for $j \in N$.*
- (3) *The dual Y is also given by $Y = \Gamma^{-1}X$ or $Y_i = \sum_{j \in N} \gamma^{i,j} X_j$, $i \in N$, where $\Gamma^{-1} = (\gamma^{i,j})_{i,j \in N}$.*
- (4) *The covariance matrix of Y is equal to Γ^{-1} .*
- (5) *The dual of Y is X .*
- (6) *$\text{sp}\{X_j; j \in N\} = \text{sp}\{Y_j; j \in N\}$.*

Proof. First, we prove (1). Let Z be another dual of X and $j \in N$ be fixed. Then $(X_i, Y_j - Z_j) = 0$ for all $i \in N$. However, since $Y_j - Z_j \in \text{sp}\{X_k; k \in N\}$, it follows that $Y_j = Z_j$ and hence (1). (2) follows from the proof of Proposition 2.2. To prove (3) and (4), we put $Y = \Gamma^{-1}X$. Then $Y_j \in \text{sp}\{X_k; k \in N\}$. Since Γ^{-1} is Hermitian, we have

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(X, X) \Gamma^{-1} = \Gamma \Gamma^{-1} = I, \\ \text{Cov}(Y, Y) &= \Gamma^{-1} \text{Cov}(X, X) \Gamma^{-1} = \Gamma^{-1} \Gamma \Gamma^{-1} = \Gamma^{-1}. \end{aligned}$$

Thus (3) and (4) follow. Finally, we obtain (5) and (6) from (3) and (4). \square

From the two representations in Proposition 2.3 (2), (3) for the dual Y , we find the following representation for the standardized interpolation error:

$$\frac{X_i - \hat{X}_i(N_i)}{\|X_i - \hat{X}_i(N_i)\|^2} = \sum_{j \in N} \gamma^{i,j} X_j \quad \text{with } N_i = N \setminus \{i\}.$$

In particular, $\gamma^{i,i} = 1/\|X_i - \hat{X}_i(N_i)\|^2$. Notice that these equalities hold even if Γ is not a Toeplitz matrix or X is not a segment of a stationary process. For some statistical/physical interpretations of the entries of Γ^{-1} , the inverse of a stationary covariance matrix, see Bhansali (1990) and references therein.

Now, we are ready to state the main duality lemma.

Lemma 2.4. *Let N be a finite index set. Assume that $X \in H_N$ has the dual $Y \in H_N$ and that K , M and a singleton $\{l\}$ partition N , i.e.,*

$$N = K \cup \{l\} \cup M \quad (\text{disjoint union}).$$

Then the following equalities hold:

- (a) $X_l - \hat{X}_l(K) = \frac{Y_l - \hat{Y}_l(M)}{\|Y_l - \hat{Y}_l(M)\|^2}$.
- (b) $\|X_l - \hat{X}_l(K)\| = \frac{1}{\|Y_l - \hat{Y}_l(M)\|}$.

Proof. Since X and Y are minimal and biorthogonal, $X_l - \hat{X}_l(K)$ and $Y_l - \hat{Y}_l(M)$ are nonzero and belong to the same one-dimensional space, that is, the orthogonal

complement of $\text{sp}\{X_j; j \in K\} \oplus \text{sp}\{Y_j; j \in M\}$ in $\text{sp}\{X_j; j \in N\}$. Therefore, one is a multiple of the other; for some $c \in \mathbb{C}$,

$$X_l - \hat{X}_l(K) = c \frac{Y_l - \hat{Y}_l(M)}{\|Y_l - \hat{Y}_l(M)\|^2}.$$

But, since c is equal to

$$c \frac{(Y_l - \hat{Y}_l(M), Y_l - \hat{Y}_l(M))}{\|Y_l - \hat{Y}_l(M)\|^2} = c \frac{(Y_l - \hat{Y}_l(M), Y_l)}{\|Y_l - \hat{Y}_l(M)\|^2} = (X_l - \hat{X}_l(K), Y_l) = (X_l, Y_l) = 1,$$

we get (a) and (b) and hence the lemma. \square

In the applications in Section 3, we use this duality in the form of the next lemma which gives a way of computing the predictor coefficients and prediction error variance using the inverse matrix $\Gamma^{-1} = (\gamma^{i,j})$.

Lemma 2.5. *Let $N, X = (X_j)_{j \in N}, Y = (Y_j)_{j \in N}, K, M$ and $\{l\}$ be as in Lemma 2.4 with $\Gamma = (\gamma_{i,j})_{i,j \in N}$ the covariance matrix of X and $\Gamma^{-1} = (\gamma^{i,j})_{i,j \in N}$. Then*

$$(2.1) \quad X_l - \hat{X}_l(K) = \sum_{i \in M \cup \{l\}} \alpha'_i Y_i,$$

$$(2.2) \quad \|X_l - \hat{X}_l(K)\|^2 = \alpha'_l,$$

where $(\alpha'_i)_{i \in M \cup \{l\}}$ is the solution to the following system of linear equations:

$$(2.3) \quad \sum_{i \in M \cup \{l\}} \alpha'_i \gamma^{i,j} = \delta_{lj}, \quad j \in M \cup \{l\}.$$

In particular, the prediction error variance $\sigma_l^2(K) = \|X_l - \hat{X}_l(K)\|^2$ is given by

$$(2.4) \quad \sigma_l^2(K) = \text{the } (l,l)\text{-entry of the inverse of } (\gamma^{i,j})_{i,j \in M \cup \{l\}},$$

and the predictor coefficients α_k in $\hat{X}_l(K) = \sum_{k \in K} \alpha_k X_k$ are given by

$$(2.5) \quad \alpha_k = - \sum_{i \in M \cup \{l\}} \alpha'_i \gamma^{i,k}, \quad k \in K,$$

whence we have

$$(2.6) \quad \hat{X}_l(K) = - \sum_{k \in K} \left(\sum_{i \in M \cup \{l\}} \alpha'_i \gamma^{i,k} \right) X_k.$$

Proof. Since Y_j 's are linearly independent, Lemma 2.4 (a) shows that $X_l - \hat{X}_l(K)$ is uniquely expressed in the form (2.1). Then $\alpha'_l = \|Y_l - \hat{Y}_l(M)\|^{-2}$, which, in view of Lemma 2.4 (b), is equal to $\|X_l - \hat{X}_l(K)\|^2$, and (2.2) holds. Since $(X_i, Y_j) = \delta_{ij}$ and $(Y_i, Y_j) = \gamma^{i,j}$, the predictor coefficients α_k in $\hat{X}_l(K) = \sum_{k \in K} \alpha_k X_k$ satisfy

$$\alpha_k = (\hat{X}_l(K), Y_k) = \left(X_l - \sum_{i \in M \cup \{l\}} \alpha'_i Y_i, Y_k \right) = - \sum_{i \in M \cup \{l\}} \alpha'_i \gamma^{i,k}.$$

Thus (2.5), whence (2.6). Similarly, for $j \in M \cup \{l\}$, we have $(\hat{X}_l(K), Y_j) = 0$ and

$$\sum_{i \in M \cup \{l\}} \alpha'_i \gamma^{i,j} = \sum_{i \in M \cup \{l\}} \alpha'_i (Y_i, Y_j) = (X_l - \hat{X}_l(K), Y_j) = \delta_{lj}.$$

Therefore, (2.3) follows. Finally, we obtain (2.4) from (2.2) and (2.3). \square

Recall that the predictor coefficients $\alpha_k = \alpha_{k,l}(K)$ in $\hat{X}_l(K) = \sum_{k \in K} \alpha_k X_k$ and the prediction error variance $\sigma^2 = \sigma_l^2(K) = \|X_l - \hat{X}_l(K)\|^2$ are traditionally computed from $(\gamma_{i,j})_{i,j \in K \cup \{l\}}$ by solving the normal equations:

$$(2.7) \quad \begin{cases} \sum_{k \in K} \alpha_k \gamma_{k,j} = \gamma_{l,j}, & j \in K, \\ \sigma^2 = \gamma_{l,l} - \sum_{k \in K} \alpha_k \gamma_{k,l}. \end{cases}$$

Alternatively, one could write the above as an analogue of the Yule–Walker equations:

$$(2.8) \quad \gamma_{l,j} - \sum_{k \in K} \alpha_k \gamma_{k,j} = \delta_{lj} \sigma^2, \quad j \in K \cup \{l\}.$$

Then $\sigma^2 = \sigma_l^2(K)$ can be identified as

$$(2.9) \quad \sigma_l^2(K) = [\text{the } (l, l)\text{-entry of the inverse of } (\gamma_{i,j})_{i,j \in K \cup \{l\}}]^{-1}.$$

In addition, using the Cramer's rule, one may write σ^2 in (2.9) as the ratio of the two relevant determinants: $\sigma^2 = \det(\gamma_{i,j})_{i,j \in K \cup \{l\}} / \det(\gamma_{i,j})_{i,j \in K}$.

In spite of the simplicity of (2.7)–(2.9), they are not convenient for the study of the asymptotic behaviors of the predictor coefficients and predictor variance as K gets large. The method of computation in Lemma 2.5 becomes particularly useful when K is large but M is small (see Section 3.2 below).

3. APPLICATIONS TO PREDICTION PROBLEMS

In this section, we illustrate the role of the finite duality principle (Lemmas 2.4 and 2.5) in unifying some diverse prediction problems for a zero-mean, weakly stationary process $\{X_j\}_{j \in \mathbb{Z}}$ with the autocovariance function $\gamma = \{\gamma_j\}_{j \in \mathbb{Z}}: \gamma_{i-j} = (X_i, X_j)$.

For simplicity, we assume that $\{X_j\}_{j \in \mathbb{Z}}$ is purely nondeterministic, so it admits the MA representation (Wold decomposition)

$$(3.1) \quad X_j = \sum_{k=-\infty}^j b_{j-k} \varepsilon_k, \quad j \in \mathbb{Z},$$

where $\{\varepsilon_j\}_{j \in \mathbb{Z}}$ is the normalized innovation of $\{X_j\}_{j \in \mathbb{Z}}$ defined by

$$\varepsilon_j := \{X_j - \hat{X}_j(\{\dots, j-2, , j-1\})\} / \|X_j - \hat{X}_j(\{\dots, j-2, , j-1\})\|, \quad j \in \mathbb{Z},$$

and $\{b_k\}_{k=0}^\infty$ is the MA coefficients given by $b_k := (X_0, \varepsilon_{-k})$. We define a sequence of complex numbers $\{a_k\}_{k=0}^\infty$ by the relation

$$(3.2) \quad \sum_{k=0}^j b_k a_{j-k} = \delta_{0j}, \quad j \geq 0.$$

If the series $\sum_{j=0}^\infty a_j X_{-j}$ is mean-convergent, then (3.1) can be inverted as

$$(3.3) \quad \varepsilon_j = \sum_{k=-\infty}^j a_{j-k} X_k, \quad j \in \mathbb{Z}.$$

This is essentially the same as the AR representation (see Pourahmadi (2001)), and we call $\{a_k\}$ the AR coefficients of $\{X_j\}_{j \in \mathbb{Z}}$. As suggested in (1.2) and (1.4), these $\{b_k\}$ and $\{a_k\}$ play an important role in prediction problems.

3.1. Finite Prediction Problems with Missing Values. Let M be a finite set of integers that does not contain zero. Throughout this section, it represents the index set of missing (unknown) values when predicting X_0 . For given M , we take the integers $m, n \geq 0$ so large that $M \subset N := \{-m, \dots, n\}$, and put $K = N \setminus (M \cup \{0\})$, which represents the index set of the observed values, so that we have the partition $N = K \cup \{0\} \cup M$ as in Lemma 2.4. We start with the prediction problem for a finite index set K . Once the problem is solved for such a K , the solutions for infinite index sets $S_n \setminus M$ and $S_\infty \setminus M$ are obtained by taking the limit of the solutions, first as $m \rightarrow \infty$, and then $n \rightarrow \infty$.

Traditionally, the coefficients of the finite linear predictor $\hat{X}_0(K)$ and its prediction error variance $\sigma^2(K) = \|X_0 - \hat{X}_0(K)\|^2$ are expressed in terms of the covariance function γ , using the normal equations (2.7). However, the results so obtained are not convenient for studying the asymptotic behaviors of the predictor coefficients as $m \rightarrow \infty$ and/or $n \rightarrow \infty$. The problem can be made much simpler by the finite duality principle and some fundamental facts about the finite MA and AR representations, as we explain now (see also Pourahmadi (2001)).

For the future segment $\{X_j\}_{j=0}^\infty$ of the process, we define its normalized innovation $\{\varepsilon_{j,0}\}_{j=0}^\infty$ by the Gram–Schmidt method: $\varepsilon_{0,0} := X_0/\|X_0\|$ and

$$\varepsilon_{j,0} := \{X_j - \hat{X}_j(\{0, \dots, j-1\})\}/\|X_j - \hat{X}_j(\{0, \dots, j-1\})\|, \quad j \geq 1.$$

Then $\{X_j\}$ and $\{\varepsilon_{j,0}\}$ admit the following finite MA and AR representations:

$$X_j = \sum_{k=0}^j b_{j-k,j} \varepsilon_{k,0}, \quad \varepsilon_{j,0} = \sum_{k=0}^j a_{j-k,j} X_k, \quad j \geq 0.$$

Here $\{b_{k,j}\}_{k=0}^j$ is defined by $b_{k,j} := (X_j, \varepsilon_{j-k,0})$ and $\{a_{k,j}\}_{k=0}^j$ by

$$\sum_{k=i}^j b_{j-k,j} a_{k-i,k} = \delta_{ij} \quad \text{or} \quad \sum_{k=i}^j a_{j-k,j} b_{k-i,k} = \delta_{ij}, \quad i \leq j.$$

These finite MA and AR coefficients converge to their infinite counterparts:

$$(3.4) \quad \lim_{j \rightarrow \infty} b_{k,j} = b_k, \quad \lim_{j \rightarrow \infty} a_{k,j} = a_k.$$

If we consider $\{X_j\}_{j=-m}^\infty$ instead of $\{X_j\}_{j=0}^\infty$, then by stationarity, it follows that

$$(3.5) \quad X_j = \sum_{k=-m}^j b_{j-k,m+j} \varepsilon_{k,-m}, \quad \varepsilon_{j,-m} = \sum_{k=-m}^j a_{j-k,m+j} X_k, \quad j \geq -m,$$

where $\{\varepsilon_{j,-m}\}_{j=-m}^\infty$ is the normalized innovation of $\{X_j\}_{j=-m}^\infty$ defined in the same way. We notice that

$$(3.6) \quad \varepsilon_j = \lim_{m \rightarrow \infty} \varepsilon_{j,-m}, \quad j \in \mathbb{Z}.$$

Thus, the representations in (3.5) reduce to (3.1) and (3.3) as $m \rightarrow \infty$.

Recall that $N = \{-m, \dots, n\}$ and let X be the vector $(X_j)_{j \in N}$ with covariance matrix $\Gamma = (\gamma_{i-j})_{i,j \in N}$. From Proposition 2.3 (3), its dual Y is given by $Y = \Gamma^{-1}X$. Let ε be the normalized innovation vector of X , i.e., $\varepsilon := (\varepsilon_{j,-m})_{j \in N}$. Then it follows from (3.5) that

$$X = B\varepsilon, \quad \varepsilon = AX,$$

where A and B are the lower triangular matrices with (i,j) -entries $a_{i-j,m+i}$ and $b_{i-j,m+i}$ for $-m \leq j \leq i \leq n$, respectively. Since $A = B^{-1}$ and $\Gamma = BB^*$, we have

$$\Gamma^{-1} = A^*A, \quad Y = A^*\varepsilon.$$

Thus, the (i, j) -entry $\gamma^{i,j}$ of Γ^{-1} and the j -th entry Y_j of Y have the representations

$$(3.7) \quad \gamma^{i,j} = \sum_{k=i \vee j}^n \bar{a}_{k-i,m+k} a_{k-j,m+k}, \quad Y_j = \sum_{k=j}^n \bar{a}_{k-j,m+k} \varepsilon_{k,-m},$$

which are certainly more conducive to studying their limits as first $m \rightarrow \infty$ and then $n \rightarrow \infty$, see (3.4) and (3.6).

Now, we are ready to express the predictor $\hat{X}_0(K)$, the prediction error $X_0 - \hat{X}_0(K)$ and its variance $\sigma^2(K)$ as prescribed by Lemma 2.5. In particular, it follows from (2.1), (2.4) and (3.7) that

$$(3.8) \quad \sigma^2(K) = \text{the } (0,0)\text{-entry of the inverse of } \left(\sum_{k=i \vee j}^n \bar{a}_{k-i,m+k} a_{k-j,m+k} \right)_{i,j \in M \cup \{0\}}$$

and

$$(3.9) \quad X_0 - \hat{X}_0(K) = \sum_{i \in M \cup \{0\}} \alpha'_i \left(\sum_{k=i}^n \bar{a}_{k-i,m+k} \varepsilon_{k,-m} \right),$$

where α'_i 's are as in Lemma 2.5 with $l = 0$.

To highlight some far-reaching consequences of (3.8) and (3.9), a few special cases corresponding to the classical prediction problems of Kolmogorov (1941), Yaglom (1963) and Nakazi (1984) are singled out and listed as examples in the next section according to the cardinality of the index set M of the missing values.

3.2. Examples. In this section, we discuss three distinct examples of the use of the finite duality principle and illustrate the process of obtaining results for the two infinite index sets $S = S_n \setminus M$ and $S = S_\infty \setminus M$.

Since $\{X_j\}_{j \in \mathbb{Z}}$ is purely nondeterministic, it has the spectral density function f with $\log f \in L^1$: $\gamma_j = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-ij\lambda} f(\lambda) d\lambda$. Also, there exists an outer function h in the Hardy class H^2 such that $f = |h|^2$ and $h(0) > 0$, and we have

$$(3.10) \quad h(z) = \sum_{k=0}^{\infty} b_k z^k, \quad \frac{1}{h(z)} = \sum_{k=0}^{\infty} a_k z^k$$

in the unit disc. This shows that $f^{-1} \in L^1$ if and only if $\{a_k\}$ is square summable. Using (3.10), which should be compared with (3.1)–(3.3), we can define the MA and AR coefficients in an analytical way.

Example 3.1 (The Finite Kolmogorov–Nakazi Problem). This is a finite interpolation problem corresponding to $K = \{-m, \dots, n\} \setminus \{0\}$ and $M = \emptyset$ (empty set), and the solution of (2.3) is $\alpha'_0 = 1/\gamma^{0,0}$. Consequently, from (3.7)–(3.9), we have

$$(3.11) \quad \sigma^2(K) = \left(\sum_{k=0}^n |a_{k,m+k}|^2 \right)^{-1}$$

and

$$(3.12) \quad X_0 - \hat{X}_0(K) = \left(\sum_{k=0}^n |a_{k,m+k}|^2 \right)^{-1} \sum_{k=0}^n \bar{a}_{k,m+k} \varepsilon_{k,-m}.$$

Next, we show that (3.11) and (3.12) are precursors of important results in prediction theory due to Kolmogorov (1941), Masani (1960), and Nakazi (1984).

The result (1.4) of Nakazi (1984) for $S_n = \{\dots, n-1, n\} \setminus \{0\}$ is obtained by taking the limit of (3.11) as $m \rightarrow \infty$ (without assuming $f^{-1} \in L^1$). Indeed, by (3.4), we see that (3.11) gives

$$(3.13) \quad \sigma^2(S_n) = \left(\sum_{k=0}^n |a_k|^2 \right)^{-1}.$$

Also, in view of (3.6), it follows from (3.12) that

$$(3.14) \quad X_0 - \hat{X}_0(S_n) = \left(\sum_{k=0}^n |a_k|^2 \right)^{-1} \sum_{k=0}^n \bar{a}_k \varepsilon_k.$$

The solution (1.3) of the Kolmogorov (1941) interpolation problem with $S_\infty = \mathbb{Z} \setminus \{0\}$ follows from (3.13) by taking the limit as $n \rightarrow \infty$, provided that $\{a_k\}$ is square summable. Thus, as in Kolmogorov (1941), assuming that $\{X_t\}$ is minimal or $f^{-1} \in L^1$, we obtain

$$\sigma^2(S_\infty) = \left(\sum_{k=0}^\infty |a_k|^2 \right)^{-1} = \left(\frac{1}{2\pi} \int_{-\pi}^\pi f(\lambda)^{-1} d\lambda \right)^{-1}.$$

Under the same minimality condition, the limit of (3.14) as $n \rightarrow \infty$, leads to

$$X_0 - \hat{X}_0(S_\infty) = \left(\sum_{k=0}^\infty |a_k|^2 \right)^{-1} \sum_{k=0}^\infty \bar{a}_k \varepsilon_k,$$

which is Masani's (1960) representation of the two-sided innovation of $\{X_j\}$ at time 0. It is instructive to note that this is a moving average in terms of the future innovations. In fact, the source of such moving average representation can be traced to (3.7) and (3.14). A version of (3.14) seems to have appeared first in Box and Tiao (1975) in the context of intervention analysis; see Pourahmadi (1989), and Pourahmadi (2001, Section 8.4) for a more rigorous derivation, detailed discussion and connection with outlier detection.

Our second example corresponds to M having cardinality one and hence involves inversion of 2×2 matrices, no matter how large K is.

Example 3.2 (The Finite Past with a Single Missing Value). This problem corresponds to $m > 0$, $n = 0$, $K = \{-m, \dots, -1\} \setminus \{-u\}$ and $M = \{-u\}$, where $1 \leq u \leq m$, so that X_{-u} from the finite past of length m is missing. By (3.7), the 2×2 matrix for solving (2.3) is

$$\begin{pmatrix} \gamma^{-u,-u} & \gamma^{-u,0} \\ \gamma^{0,-u} & \gamma^{0,0} \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^u |a_{u-k,m-k}|^2 & a_{0,m} \bar{a}_{u,m} \\ \bar{a}_{0,m} a_{u,m} & |a_{0,m}|^2 \end{pmatrix}.$$

Hence, using the subscript m to emphasize the dependence on m , we have

$$\alpha'_{0,m} = \frac{1}{\Delta_m} \sum_{k=0}^u |a_{u-k,m-k}|^2, \quad \alpha'_{-u,m} = -\frac{\bar{a}_{0,m} a_{u,m}}{\Delta_m},$$

with the determinant $\Delta_m = |a_{0,m}|^2 \sum_{k=1}^u |a_{u-k,m-k}|^2$. Thus, by (3.8) and (3.9),

$$(3.15) \quad \begin{cases} \sigma^2(K) = \frac{\sum_{k=0}^u |a_{u-k,m-k}|^2}{|a_{0,m}|^2 \sum_{k=1}^u |a_{u-k,m-k}|^2}, \\ X_0 - \hat{X}_0(K) = \alpha'_{0,m} \bar{a}_{0,m} \varepsilon_{0,-m} + \alpha'_{-u,m} \sum_{k=0}^u \bar{a}_{u-k,m-k} \varepsilon_{-k,-m}, \end{cases}$$

and, taking the limit as $m \rightarrow \infty$,

$$(3.16) \quad \begin{cases} \sigma^2(S_0 \setminus \{-u\}) = |b_0|^2 \frac{\sum_{k=0}^u |a_k|^2}{\sum_{k=0}^{u-1} |a_k|^2}, \\ X_0 - \hat{X}_0(S_0 \setminus \{-u\}) = \alpha'_0 \bar{a}_0 \varepsilon_0 + \alpha'_{-u} \sum_{k=0}^u \bar{a}_{u-k} \varepsilon_{-k}, \end{cases}$$

where α'_0 and α'_{-u} are the limits of $\alpha'_{0,m}$ and $\alpha'_{-u,m}$, as $m \rightarrow \infty$, respectively.

The expressions in (3.16) were obtained first in Pourahmadi (1992); see also Pourahmadi and Soofi (2000) and Pourahmadi (2001, Section 8.3). However, those

in (3.15) have not appeared before. For $n > 0$, slightly more general calculations leading to analogues of (3.15) and (3.16) can be used to show that the inverse autocorrelation function of $\{X_t\}$ at lag u is the negative of the partial correlation between X_0 and X_u after elimination of the effects of X_t , $t \neq 0, u$, as shown in Kanto (1984) for processes with strictly positive spectral density functions.

Example 3.3 (The Finite Yaglom Problem). There are many situations where the cardinality of M is two or more; see Pourahmadi *et al.* (2007), Box and Tiao (1975), Brubacher and Wilson (1976), Damsleth (1980), Abraham (1981). In the literature of time series analysis, there are several ad hoc methods for interpolating the missing values. For example, Brubacher and Wilson (1976) minimize

$$\sum_{-m}^n \varepsilon_j^2 = \sum_{-m}^n \left(\sum_{k=-\infty}^j a_{j-k} X_k \right)^2$$

with respect to the unknown X_j , $j \in M \cup \{0\}$, and then study the solution of the normal equations as $m, n \rightarrow \infty$. Budinsky (1989) has shown that this approach under some conditions gives the same result as the more rigorous approach of Yaglom (1963). In applying Lemma 2.5 to this problem, we first note that, due to the large cardinality of M , handling (3.8) and (3.9) via (2.3) does not lead to simple explicit formulas as in (3.15) and (3.16). Nevertheless, the limits of the expressions in (3.8) and (3.9) as first $m \rightarrow \infty$, and then as $n \rightarrow \infty$ (assuming $f^{-1} \in L^1$) have simple forms in terms of the AR parameters:

$$(3.17) \quad \begin{cases} X_0 - \hat{X}_0(S) = \sum_{i \in M \cup \{0\}} \alpha'_i \left(\sum_{k=i}^{\infty} \bar{a}_{k-i} \varepsilon_k \right), \\ \sigma^2(S) = \text{the } (0, 0)\text{-entry of the inverse of } \left(\sum_{k=i \vee j}^{\infty} \bar{a}_{k-i} a_{k-j} \right)_{i, j \in M \cup \{0\}}. \end{cases}$$

Now, using (3.10) and writing the entries of the above matrix, in terms of the Fourier coefficients of f^{-1} , it follows that (3.17) reduces to the results in Yaglom (1963); see also Salehi (1979).

3.3. The Infinite Past and the Wold Decomposition. A more direct method of solving prediction problems for $S = S_n \setminus M$ is to reduce them to a different class of finite prediction problems than those in Section 3.2. This is done by using the Wold decomposition of a purely nondeterministic stationary process.

As in Section 3.1, write $N = \{-m, \dots, n\}$ and $N = K \cup \{0\} \cup M$ (disjoint), so that $S = S_n \setminus M = \{\dots, -m-2, -m-1\} \cup K$ (disjoint). For $j \geq -m$, let \hat{X}_j be the linear least-squares predictor of X_j based on the infinite past $\{X_k; k < -m\}$. Then, by (3.1),

$$X_j - \hat{X}_j = \sum_{k=-m}^j b_{j-k} \varepsilon_k, \quad j \geq -m,$$

which are orthogonal to $\overline{\text{sp}}\{X_j; j < -m\}$, and it follows that

$$\overline{\text{sp}}\{X_j; j \in S\} = \text{sp}\{X_j - \hat{X}_j; j \in K\} \oplus \overline{\text{sp}}\{X_j; j < -m\}.$$

This equality plays the key role in finding the predictor of X_0 and its prediction error variance, based on $\{X_j; j \in S\}$. In fact, by using it, we only have to solve the problem of predicting $X_0 - \hat{X}_0$ based on $\{X_j - \hat{X}_j; j \in K\}$. More precisely, we consider $X' := (X_j - \hat{X}_j)_{j \in N}$ which has the covariance matrix $G = (g_{i,j})_{i,j \in N}$ with

$$g_{i,j} := \sum_{k=-m}^{i \wedge j} b_{i-k} \bar{b}_{j-k}$$

(see Pourahmadi (2001, p. 273)). Then, writing $X_0 = \hat{X}_0 + (X_0 - \hat{X}_0)$, we get

$$(3.18) \quad \begin{aligned} \hat{X}_0(S) &= \hat{X}_0 + \sum_{k \in K} \alpha_k (X_k - \hat{X}_k), \\ \sigma^2(S) &= \left\| (X_0 - \hat{X}_0) - \sum_{k \in K} \alpha_k (X_k - \hat{X}_k) \right\|^2, \end{aligned}$$

where $\sum_{k \in K} \alpha_k (X_k - \hat{X}_k)$ is the predictor of $X_0 - \hat{X}_0$ based on $\{X_k - \hat{X}_k; k \in K\}$, and the predictor coefficients α_k and prediction error variance $\sigma^2(S)$ are obtained from the normal equations (2.7) with $\gamma_{i,j}$ replaced by $g_{i,j}$; in particular, by (2.9),

$$\sigma^2(S) = \left[\text{the } (0,0)\text{-entry of the inverse of } \left(\sum_{k=-m}^{i \wedge j} b_{i-k} \bar{b}_{j-k} \right)_{i,j \in K \cup \{0\}} \right]^{-1}.$$

We can also apply Lemma 2.5 to the above finite prediction problem for X' . In so doing, the following representations for the (i,j) -entry $g^{i,j}$ of G^{-1} and the j -th entry Y_j of the dual Y of X' are available:

$$(3.19) \quad g^{i,j} = \sum_{k=i \vee j}^n \bar{a}_{k-i} a_{k-j}, \quad Y_j = \sum_{k=j}^n \bar{a}_{k-j} \varepsilon_k.$$

In fact, these are obtained by using (3.2) and Proposition 2.3 (3) or by letting $m \rightarrow \infty$ in (3.7). The explicit representations in (3.19) are also important in finding series representations for predictors and interpolators discussed in the next two subsections.

3.4. Series Representation of the Predictors. The Wold decomposition (3.1) is often used to express predictors and prediction errors in terms of the innovation process $\{\varepsilon_t\}$. This strategy works well for achieving the goal (P₂) in Section 1, but since the innovation ε_t is not directly observable the resulting predictor formulas are not suitable for computation. To get around this difficulty, one must express the innovations or the predictors in terms of the past observations. In this section, we obtain series representations for the infinite past predictors in terms of the observed values. A novelty of our approach is its reliance on the representation of the prediction error in terms of the dual Y in (3.19), hence the solution of the problem (P₁) for $S = S_n \setminus M$ is more direct and simpler than the procedures of Bondon (2002, Theorem 3.1) and Nikfar (2006).

Assuming that $\{X_j\}_{j \in \mathbb{Z}}$ has the mean-convergent AR representation (3.3), it follows from (3.18) with $S = \{\dots, -m-2, -m-1\} \cup K$ that

$$\hat{X}_0(S) = \sum_{k \in K} \alpha_k X_k + \sum_{j=1}^{\infty} \left(f_{j,m} - \sum_{k \in K} \alpha_k f_{j,m+k} \right) X_{-m-j},$$

where $f_{j,k} := -\sum_{i=0}^k b_{k-i} a_{j+i}$ is the coefficient of the $(k+1)$ -step ahead predictor based on the infinite past $S_0 = \{\dots, -2, -1\}$, i.e., $\hat{X}_k(S_0) = \sum_{j=1}^{\infty} f_{j,k} X_{-j}$ for $k = 0, 1, \dots$. On the other hand, from the finite duality principle or, more precisely, (2.1) with (3.19), we have

$$\hat{X}_0(S) = X_0 - \sum_{i \in M \cup \{0\}} \alpha'_i \left(\sum_{k=i}^n \bar{a}_{k-i} \varepsilon_k \right).$$

From this, replacing ε_k from (3.3) and after some algebra, we get the following alternative series representation for the predictor of X_0 based on the incomplete past:

$$(3.20) \quad \hat{X}_0(S) = - \sum_{j \in S} \left(\sum_{i \in M \cup \{0\}} \alpha'_i \sum_{k=i \vee j}^n \bar{a}_{k-i} a_{k-j} \right) X_j.$$

We note that the prediction error here has the representation

$$(3.21) \quad X_0 - \hat{X}_0(S) = \sum_{i \in M \cup \{0\}} \alpha'_i \left(\sum_{k=i}^n \bar{a}_{k-i} \varepsilon_k \right)$$

in terms of the dual Y in (3.19). Furthermore, the sequence $\{\sum_{k=j}^n \bar{a}_{k-j} \varepsilon_k\}_{j=-\infty}^n$ spans $\overline{\text{sp}}\{X_j; j \leq n\}$, the infinite past up to n of the process $\{X_t\}$. The formulas (3.20) and (3.21) were obtained initially by Bondon (2002, Theorem 3.2) without using the notion of duality.

3.5. Series Representation of the Interpolators. Series representation for the interpolator of X_0 based on the observed values from the index set $S = S_\infty \setminus M = \mathbb{Z} \setminus (M \cup \{0\})$ was obtained by Salehi (1979). Here we obtain such representation using the idea of the dual process. Assuming $f^{-1} \in L^1$ or $\sum_{j=0}^\infty |a_j|^2 < \infty$, the process

$$\xi_j := \sum_{k=j}^\infty \bar{a}_{k-j} \varepsilon_k, \quad j \in \mathbb{Z},$$

is well-defined in the sense of mean-square convergence. From (3.1), (3.2), and the above results, we have the following:

- (i) $(X_i, \xi_j) = \delta_{ij}$ for $i, j \in \mathbb{Z}$.
- (ii) $\xi_j = \{X_j - \hat{X}_j(\mathbb{Z} \setminus \{j\})\} / \|X_j - \hat{X}_j(\mathbb{Z} \setminus \{j\})\|^2$ for $j \in \mathbb{Z}$.
- (iii) $\{\xi_j; j \in \mathbb{Z}\}$ spans the space $\overline{\text{sp}}\{X_j; j \in \mathbb{Z}\}$.
- (iv) $\{\xi_j; j \in \mathbb{Z}\}$ is a stationary process with the autocovariance function

$$\gamma^j := \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ij\lambda} f(\lambda)^{-1} d\lambda, \quad j \in \mathbb{Z},$$

$$\text{i.e., } (\xi_i, \xi_j) = \gamma^{i-j} = \sum_{k=i \vee j}^\infty \bar{a}_{k-i} a_{k-j} \text{ for } i, j \in \mathbb{Z}.$$

The process $\{\xi_j\}$ has already appeared in prediction theory and time series analysis, and is called the *standardized two-sided innovation* (Masani (1960)) or the *inverse process* (Cleveland (1972)) of $\{X_t\}_{t \in \mathbb{Z}}$.

Now, for solving the interpolation problem with $S = \mathbb{Z} \setminus (M \cup \{0\})$, we need to show that $\{\xi_j; j \in M \cup \{0\}\}$ spans the orthogonal complement of $\overline{\text{sp}}\{X_j; j \in S\}$ in $\overline{\text{sp}}\{X_j; j \in \mathbb{Z}\}$. Then, it turns out that there is unique $(\alpha'_j)_{j \in M \cup \{0\}}$ satisfying

$$X_0 - \hat{X}_0(S) = \sum_{i \in M \cup \{0\}} \alpha'_i \xi_i = \sum_{i \in M \cup \{0\}} \alpha'_i \left(\sum_{k=i}^\infty \bar{a}_{k-i} \varepsilon_k \right)$$

(see (3.9) and (3.21)), and that $\sigma^2(S) = \alpha'_0$. Since

$$(X_0, \xi_j) - \sum_{i \in M \cup \{0\}} \alpha'_i (\xi_i, \xi_j) = (\hat{X}_0(S), \xi_j) = 0, \quad j \in M \cup \{0\},$$

we can compute $(\alpha'_i)_{i \in M \cup \{0\}}$ by solving the following system of linear equations:

$$\sum_{i \in M \cup \{0\}} \alpha'_i \gamma^{i-j} = \delta_{j0}, \quad j \in M \cup \{0\}.$$

As for the predictor, if $\sum_{j=-\infty}^\infty \gamma^j X_{-j}$ is mean-convergent, then $(\xi_j)_{j \in \mathbb{Z}}$ admits the representation

$$\xi_i = \sum_{j=-\infty}^\infty \gamma^{i-j} X_j, \quad i \in \mathbb{Z},$$

and we have

$$\hat{X}_0(S) = - \sum_{j \in S} \left(\sum_{i \in M \cup \{0\}} \alpha_i \gamma^{i-j} \right) X_j,$$

which is the two-sided version of the formula (3.20).

4. DISCUSSION AND FUTURE WORK

We have reviewed and unified some important results from prediction theory of stationary processes using a finite-dimensional duality principle whose proof is based on elementary ideas from the linear algebra. Our time-domain, geometric and finite-dimensional approach brings considerable clarity and simplicity to this area of prediction theory as compared to the classical spectral-domain approach based on analytic function theory and duality in the infinite-dimensional spaces. Since our duality lemma is not confined to stationary processes or Toeplitz matrices, it has the potential of being useful in solving similar prediction problems for nonstationary processes, particularly those with low displacement ranks (Kailath and Sayed (1995)). However, the present form of the lemma does not seem to be useful for prediction problems of infinite-variance or L^p -processes (Cambanis and Soltani (1984), Cheng *et al.* (1998)).

REFERENCES

- [1] AGRAHAM, B. (1981). Missing observations in time series. *Commun. Statist.-Theor. Meth.* **A10**, 1643-1653.
- [2] BHANSALI, R.J. (1990). On a relationship between the inverse of a stationary covarinace matrix and the linear interpolator. *J. Appl. Prob.* **27**, 156-170.
- [3] BONDON, P. (2002). Prediction with incomplete past of a stationary process. *Stochastic Process. Appl.* **98**, 67-76.
- [4] BONDON, P. (2005). Influence of missing values on the prediction of a stationary time series. *J. of Time Series Analys.* **26**, 519-525.
- [5] BOX, G. E. P., JENKINS, G. M. AND REINSEL, G. C. (1994). *Time series analysis. Forecasting and control.* 3rd edition. Prentice Hall, Englewood Cliffs.
- [6] BOX, G.E.P. AND TIAO, G.C. (1975). Intervention analysis with applications to environmental and economic problems. *J. of the Amer. Statist. Soc.* **70**, 70-79.
- [7] BROCKWELL, P. J. AND DAVIS, R. A. (1991). *Time series: theory and methods.* 2nd edition. Springer-Verlag, New York.
- [8] BRUBACHER, S.R. AND WILSON, T.G. (1976). Interpolating time series with application to the estimation of holiday effects on electricity demand. *Appl. Statist.* **25**, 107-116.
- [9] BUDINSKY, P. (1989). On the equivalence of two methods for interpolation. *Kybernetika*, **25**, 461-466.
- [10] CAMBANIS, S. AND SOLTANI, A. R. (1984). Prediction of stable Processes: Spectral and moving average representations. *Z. Wahrsch. Verw. Gebiete* **66**, 593-612.
- [11] CHENG, R., MIAMEE, A. G. AND POURAHMADI, M. (1998). Some extremal problems in $L^p(w)$. *Proc. Amer. Math. Soc.* **126**, 2333-2340.
- [12] CLEVELAND, W.S. (1972). The inverse autocorrelations of time series and their applications. *Technometrics*, **14**, 277-283.
- [13] DAMSLETH, E. (1980). Interpolating missing values in a time series. *Scand. J. Statist.* **7**, 33-39.
- [14] FRANK, M. AND KLOTZ, L. (2002). A duality method in prediction theory of multivariate stationary sequences. *Math. Nachr.* **244**, 64-77.
- [15] GRENANDER, U. AND ROSENBLATT, M. (1954). An extension of a theorem of G. Szegö and its applications to the study of stochastic processes. *Trans. of Amer. Math. Soc.* **76**, 112-126.
- [16] KAILATH, T. AND SAYED, A.H. (1995). Displacement structure: Theory and applications. *SIAM Review*, **37**, 297-386.
- [17] KANTO, A.J. (1984). A characterization of the inverse autocorrelation function. *Commu. Statisti.-Theor. Meth.* **13**, 2503-2510.
- [18] KLOTZ, L. AND RIEDEL, M. (2002). Some remarks on duality of stationary sequences. *Colloq. Math.* **86**, 225-228.
- [19] KOLMOGOROV, A. N. (1941). Stationary sequences in a Hilbert space. *Bull. Moscow State University* **2**, 1-40.
- [20] LANDAU, H.J. (1987). Maximum entropy and the moment problem. *Bull. of the Amer. Math. Soc.* **16**, 47-77.

- [21] MIAMEE, A. G. (1993). On basicity of exponentials in $L^p(d\mu)$ and general prediction problems. *Period. Math. Hungar.* **26**, 115–124.
- [22] MIAMEE, A. G. AND POURAHMADI, M. (1988). Best approximation in $L^p(d\mu)$ and prediction problems of Szegö, Kolmogorov, Yaglom and Nakazi. *J. London Math. Soc.* **38**, 133–145.
- [23] NAKAZI, T. (1984). Two problems in prediction theory. *Studia Math.* **78**, 7–14.
- [24] NAKAZI, T. AND TAKAHASHI, K. (1980). Prediction n units of time ahead. *Proc. Amer. Math. Soc.* **80**, 658–659.
- [25] NIKFAR, M. (2006). The generalization of Szegö's theorem. To appear.
- [26] POURAHMADI, M. (1984). Taylor expansion of $\exp(\sum_{k=0}^{\infty} a_k z^k)$ and some applications. *Amer. Math. Monthly* **91**, 303–307.
- [27] POURAHMADI, M. (1989). Estimation and interpolation of missing values of a stationary time series. *J. Time Ser. Anal.* **10**, 149–169.
- [28] POURAHMADI, M. (1992) Alternating projections and interpolation of stationary processes. *J. Appl. Probab.* **29**, 921–931.
- [29] POURAHMADI, M. (1993). Two prediction problems and extensions of a theorem of Szegö. *Bull. Iranian Math. Soc.* **19**, 1–12.
- [30] POURAHMADI, M. (2001). *Foundations of Prediction Theory and Time Series Analysis*. John Wiley, New York.
- [31] POURAHMADI, M. AND SOOFI, E. S. (2000). Prediction variance and information worth of observations in time series. *J. Time Ser. Anal.* **21**, 413–434.
- [32] POURAHMADI, M., INOUE, A. AND KASAHARA, Y. (2006). A prediction problem in $L^2(w)$. *Proc. Amer. Math. Soc.*, to appear.
- [33] POURAHMADI, M., INOUE, A. AND KASAHARA Y. (2006). A prediction problem in $L^2(w)$. Abstracts for the Japan Math. Soc. Meeting.
- [34] ROZANOV, YU. A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco.
- [35] SALEHI, H. (1979). Algorithm for linear interpolator and interpolation error for minimal stationary stochastic processes. *Ann. Probab.* **7**, 840–846.
- [36] SIMON, B. (2005). *Orthogonal Polynomials on the Unit Circle, Part 1: Classical Theory*, AMS, Providence, R.I.
- [37] URBANIK, K. (2000). A duality principle for stationary random sequences. *Colloq. Math.* **86**, 153–162.
- [38] WIENER, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications*. Cambridge, Mass; John Wiley, New York; Chapman & Hall, London
- [39] WIENER, N. AND MASANI, P. R. (1958). The prediction theory of multivariate stationary processes II. *Act. Math.* **99**, 93–137.
- [40] WOLD, H. ([1938] 1954). *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell. Stockholm.
- [41] YAGLOM, A. M. (1963). On a problem of linear interpolation of stationary random sequences and processes. *Amer. Math. Soc. Selected Translations Math. Statist.* **4**, 330–344.

DEPARTMENT OF MATHEMATICS, HOKKAIDO UNIVERSITY, SAPPORO 060-0810, JAPAN
E-mail address: y-kasa@math.sci.hokudai.ac.jp

DIVISION OF STATISTICS, NORTHERN ILLINOIS UNIVERSITY, DEKALB, IL 60115-2854, USA
E-mail address: pourahm@math.niu.edu

DEPARTMENT OF MATHEMATICS, HOKKAIDO UNIVERSITY, SAPPORO 060-0810, JAPAN
E-mail address: inoue@math.sci.hokudai.ac.jp